

日 本 国 特 許 庁  
JAPAN PATENT OFFICE

31.08.2004

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

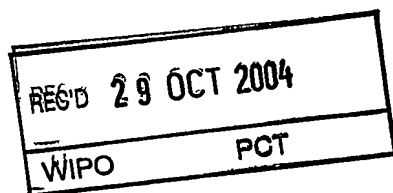
This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日  
Date of Application: 2003年 9月 5日

出 願 番 号  
Application Number: 特願2003-314247  
[ST. 10/C]: [JP2003-314247]

出 願 人  
Applicant(s): 学校法人近畿大学  
財団法人北九州産業学術推進機構  
株式会社ウエイブコム

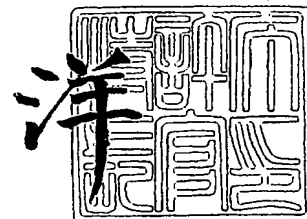
**PRIORITY DOCUMENT**  
SUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH  
RULE 17.1(a) OR (b)



2004年10月14日

特許庁長官  
Commissioner,  
Japan Patent Office

小 川



BEST AVAILABLE COPY

【書類名】 特許願  
【整理番号】 P15091751  
【あて先】 特許庁長官 今井 康夫殿  
【国際特許分類】 G10L 11/02  
【発明者】  
    【住所又は居所】 福岡県飯塚市柏の森 1 1 - 6 近畿大学 九州工学部内  
    【氏名】 五反田 博  
【発明者】  
    【住所又は居所】 福岡県飯塚市柏の森 1 1 - 6 近畿大学 九州工学部内  
    【氏名】 古屋 武志  
【発明者】  
    【住所又は居所】 福岡県飯塚市柏の森 1 1 - 6 近畿大学 九州工学部内  
    【氏名】 金田 圭市  
【特許出願人】  
    【識別番号】 000125347  
    【氏名又は名称】 学校法人近畿大学  
    【代表者】 世耕 弘昭  
【特許出願人】  
    【識別番号】 802000031  
    【氏名又は名称】 財団法人北九州産業学術推進機構  
    【代表者】 有馬 朗人  
【特許出願人】  
    【識別番号】 500331219  
    【氏名又は名称】 株式会社ウエイブコム  
    【代表者】 新井 孝弘  
【代理人】  
    【識別番号】 100090697  
    【弁理士】  
    【氏名又は名称】 中前 富士男  
【手数料の表示】  
    【予納台帳番号】 044484  
    【納付金額】 21,000円  
【提出物件の目録】  
    【物件名】 特許請求の範囲 1  
    【物件名】 明細書 1  
    【物件名】 図面 1  
    【物件名】 要約書 1

## 【書類名】特許請求の範囲

## 【請求項 1】

異なる 2 つの音源からそれぞれ発信される目的音声及び雑音を、異なる位置に設けた第 1 及び第 2 のマイクでそれぞれ受信して混合信号を形成し、該各混合信号を時間領域から周波数領域にフーリエ変換して、独立成分分析法により前記目的音声及び前記雑音にそれぞれ対応する各分割スペクトルから複数の推定スペクトル  $Y^*$ 、 $Y$  を抽出する第 1 工程と、前記各推定スペクトル  $Y^*$  を、該各推定スペクトル  $Y^*$  の振幅分布の尖度の値を分離判定基準にして前記雑音が除去された複数の推定スペクトル  $y^*$  及び前記雑音が残留する複数の推定スペクトル  $y$  に分離する第 2 工程と、前記各推定スペクトル  $y^*$  の総和  $F$  を求め、該総和  $F$  の最大値に応じて決定される閾値  $\beta$  を検出判定基準にして該総和  $F$  の周波数領域から音声区間及び雑音区間をそれぞれ検出する第 3 工程と、前記各推定スペクトル  $Y^*$  から前記音声区間に存在する成分を抽出して前記目的音声の復元スペクトル群を生成し、該復元スペクトル群を周波数領域から時間領域にフーリエ逆変換して前記目的音声を復元する第 4 工程とを有することを特徴とする定常雑音下における音声区間検出に基づく目的音声の復元方法。

## 【請求項 2】

請求項 1 記載の定常雑音下における音声区間検出に基づく目的音声の復元方法において、前記検出判定基準は、前記総和  $F$  が前記閾値  $\beta$  を超える周波数領域を前記音声区間とし、前記総和  $F$  が前記閾値  $\beta$  以下となる周波数領域を前記雑音区間として検出するように設定されていることを特徴とする定常雑音下における音声区間検出に基づく目的音声の復元方法。

## 【請求項 3】

異なる 2 つの音源からそれぞれ発信される目的音声及び雑音を、異なる位置に設けた第 1 及び第 2 のマイクでそれぞれ受信して混合信号を形成し、該各混合信号を時間領域から周波数領域にフーリエ変換して、独立成分分析法により前記目的音声及び前記雑音にそれぞれ対応する各分割スペクトルから複数の推定スペクトル  $Y^*$ 、 $Y$  を抽出する第 1 工程と、前記各推定スペクトル  $Y^*$  を、該各推定スペクトル  $Y^*$  の振幅分布の尖度の値を分離判定基準にして前記雑音が除去された複数の推定スペクトル  $y^*$  及び前記雑音が残留する複数の推定スペクトル  $y$  に分離する第 2 工程と、前記各推定スペクトル  $y^*$  の総和  $F$  を求め、該総和  $F$  の最大値に応じて決定される閾値  $\beta$  を検出判定基準にして該総和  $F$  の時間領域から音声区間及び雑音区間をそれぞれ検出する第 3 工程と、前記各推定スペクトル  $Y^*$  を周波数領域から時間領域にフーリエ逆変換して得られた前記目的音声の信号から前記音声区間に存在する成分を抽出して前記目的音声を復元する第 4 工程とを有することを特徴とする定常雑音下における音声区間検出に基づく目的音声の復元方法。

## 【請求項 4】

請求項 3 記載の定常雑音下における音声区間検出に基づく目的音声の復元方法において、前記検出判定基準は、前記総和  $F$  が前記閾値  $\beta$  を超える時間領域を前記音声区間とし、前記総和  $F$  が前記閾値  $\beta$  以下となる時間領域を前記雑音区間として検出することを特徴とする定常雑音下における音声区間検出に基づく目的音声の復元方法。

## 【請求項 5】

請求項 1～4 のいずれか 1 項に記載の定常雑音下における音声区間検出に基づく目的音声の復元方法において、前記尖度の値を前記各推定スペクトル  $Y^*$  の振幅分布のエントロピー  $H$  で評価することを特徴とする定常雑音下における音声区間検出に基づく目的音声の復元方法。

## 【請求項 6】

請求項 5 記載の定常雑音下における音声区間検出に基づく目的音声の復元方法において、前記分離判定基準は、

(1) 前記エントロピー $H$ が予め設定した閾値 $\alpha$ 未満である推定スペクトル $Y^*$ は前記推定スペクトル $y^*$ として分離し、

(2) 前記エントロピー $H$ が予め設定した閾値 $\alpha$ 以上である推定スペクトル $Y^*$ は前記推定スペクトル $y$ として分離するように設定されていることを特徴とする定常雑音下における音声区間検出に基づく目的音声の復元方法。

【書類名】明細書

【発明の名称】定常雑音下における音声区間検出に基づく目的音声の復元方法

【技術分野】

【0001】

本発明は、ICA法により得られた分離信号から音声区間と雑音区間を分離し、分離信号から音声区間に含まれる信号のみを抽出することで音声全体に含まれる残留雑音を最小化する定常雑音下における音声区間検出に基づく目的音声の復元方法に関する。

【背景技術】

【0002】

近年、音声認識技術は著しく進展して、周囲に雑音のない理想的な環境下では認識能力の極めて高い音声認識エンジンが供給できるようになってきている。

しかし、生活音や活動音が存在する一般家庭やオフィスでは、期待通りの認識率を得るのが難しい。このような環境下で音声認識エンジンのもつ本来の認識能力を引き出すには、雑音混じりの音声から雑音を除去して、話者音声のみを音声認識エンジンに受け渡す前処理が不可欠と考えられる。

その観点から、近年、独立成分分析法（以下、単にICA法という）や各種音声強調法が注目を集めており、様々なアルゴリズムが提案されている（例えば、非特許文献1～5参照）。そして、これらの各種アルゴリズムのなかで、各音源が統計的に独立という仮定の基に、音声と雑音を分離するICA法が注目されている。

【0003】

【非特許文献1】ベル、セイノフスキー（J. Bell and T. J. Sejnowski）、"アンインフォメーション マキシマイゼーション アプローチ ツー ブラインド セパレーション アンド ブラインド デイコンボルーション（An information and blind deconvolution"、ニューラル コンピューテーション（Neural Computation）、米国、エムアイティー プレス（MIT Press）、1995年6月、7巻、6号、p. 1129-1159

【非特許文献2】アマリ（S. Amari）、"ナチュラル グラディエント ワークス エフィシエントリー イン ラーニング（Natural gradient works efficiently in learning）、ニューラル コンピューテーション（Neural Computation）、米国、エムアイティー プレス（MIT Press）、1998年2月、10巻、2号、p. 254-276

【非特許文献3】リー、ギロラミ、セイノフスキー（T. W. Lee, M. Girolami and T. J. Sejnowski）、"インディペンデント コンポーネント アナリシス ユージング アン イクステンティッド インフォマックス アルゴリズム フォア ミクスト サブガウシアン アンド スーパーガウシアン ソーシズ（Independent component analysis using an extended informax algorithm for mixed subgaussian and supergaussian sources）、ニューラル コンピューテーション（Neural Computation）、米国、エムアイティー プレス（MIT Press）、1999年2月、11巻、2号、p. 417-441

【非特許文献4】ヒバリネン（A. Hyvarinen）、"ファスト アンド ロバスト フィクストポイント アルゴリズム フォア インディペンデント コンポーネント アナリシス（Fast and robust fixed-point algorithms for independent component analysis）、トランザクション ニューラル ネットワークス（IEEE Trans. Neural Networks）、米国、アイイーイーイー（IEEE）、1999年6月、10巻、3号、p. 626-634

【非特許文献5】ヒバリネン、オヤ (A. Hyvarinen, and E. Oja)、"インディペンデントコンポーネントアナリシス: アルゴリズムとアプリケーションズ (Independent component analysis: algorithms and applications)、ニューラルネットワークス (Neural Networks)、米国、ペルガモンプレス (Pergamon Press)、2000年6月、13巻、4-5号、p. 411-430

【発明の開示】

【発明が解決しようとする課題】

【0004】

しかしながら、ICA法では、残響のない理想的環境下の場合には良好に音声と雑音を分離することができるが、残響が強い実際の環境下の場合には残響に起因した残留雑音のため、分離能力が極端に低下するという問題がある。

本発明はかかる事情に鑑みてなされたもので、実際の環境下で受信した信号からICA法により作成した分離信号に基づいて音声区間と雑音区間を分離し、分離信号から音声区間に含まれる信号のみを抽出することで、目的音声に含まれる残留雑音を最小化することが可能な定常雑音下における音声区間検出に基づく目的音声の復元方法を提供することを目的とする。

【課題を解決するための手段】

【0005】

前記目的に沿う第1の発明に係る定常雑音下における音声区間検出に基づく目的音声の復元方法は、異なる2つの音源からそれぞれ発信される目的音声及び雑音を、異なる位置に設けた第1及び第2のマイクでそれぞれ受信して混合信号を形成し、該各混合信号を時間領域から周波数領域にフーリエ変換して、独立成分分析法により前記目的音声及び前記雑音にそれぞれ対応する各分割スペクトルから複数の推定スペクトル $Y^*$ 、 $Y$ を抽出する第1工程と、

前記各推定スペクトル $Y^*$ を、該各推定スペクトル $Y^*$ の振幅分布の尖度の値を分離判定基準にして前記雑音が除去された複数の推定スペクトル $y^*$ 及び前記雑音が残留する複数の推定スペクトル $y$ に分離する第2工程と、

前記各推定スペクトル $y^*$ の総和 $F$ を求め、該総和 $F$ の最大値に応じて決定される閾値 $\beta$ を検出判定基準にして該総和 $F$ の周波数領域から音声区間及び雑音区間をそれぞれ検出する第3工程と、

前記各推定スペクトル $Y^*$ から前記音声区間に存在する成分を抽出して前記目的音声の復元スペクトル群を生成し、該復元スペクトル群を周波数領域から時間領域にフーリエ逆変換して前記目的音声を復元する第4工程とを有する。

【0006】

第1及び第2のマイクでそれぞれ受信されて得られる混合信号のみを用いて2つの音源からそれぞれ発信される目的音声及び雑音を推定する場合、混合信号は畳み込み混合されている。このため、周波数領域に変換することで畳み込み混合を瞬時混合として取り扱うことができ、分離が容易となる。更に、各音源は統計的に独立と考えることができるので、独立成分分析(ICA)法を適用することができる。

ここで、ICA法を適用して得られた周波数毎の分割スペクトルには、スケーリングの不定性と成分置換の問題が生じているため、これらの問題を別途解決して目的音声及び雑音にそれぞれ対応する周波数毎の推定スペクトル $Y^*$ 、 $Y$ を抽出する。

しかし、抽出された各推定スペクトル $Y^*$ を周波数毎に見てみると、雑音が良好に分離されている周波数と雑音がほとんど除去されていない周波数が存在している状態になっている。

【0007】

一方、時間領域での音声と雑音の統計的性質として、音声は0近傍の値を取る確率が非常に大きい尖度の高い分布になるのに対して、雑音は尖度の低い分布になることが知られて

いる、このことは、周波数領域に変換した音声と雑音に対しても当てはまり、音声に対応したスペクトルの振幅分布は尖度が高く、雑音に対応したスペクトルの振幅分布は尖度が低くなる。

このことから、各推定スペクトル $Y^*$ において、その振幅分布の尖度を求めると、振幅分布の尖度が大きければその周波数では雑音が良好に除去されていると判定でき、尖度が小さければその周波数では雑音がほとんど除去されていないと判定することができる。従って、各推定スペクトル $Y^*$ の振幅分布の尖度によって、各推定スペクトル $Y^*$ を推定スペクトル $y^*$ と推定スペクトル $y$ に分離することができる。

#### 【0008】

一般に、音声の周波数成分は、時間的に変化するため、音声の特徴付けるスペクトルの現れる周波数領域は推定スペクトル $y^*$ によって異なることになる。そこで、各推定スペクトル $y^*$ の総和 $F$ を求め、周波数領域における総和 $F$ に対して総和 $F$ の最大値に応じて決定される閾値 $\beta$ を適切に選ぶことにより、時間的な音声区間をカバーするような周波数領域における音声区間を検出することができる。そして、音声区間を除いた部分を周波数領域における雑音区間として検出することができる。

従って、目的音声に対応する各推定スペクトル $Y^*$ から求めた音声区間に存在する成分のみを抽出して目的音声の復元スペクトル群を生成すると、復元スペクトル群には雑音成分は実質的に存在しないことになる。そして、この復元スペクトル群を周波数領域から時間領域にフーリエ逆変換して目的音声信号を復元することができる。

#### 【0009】

第1の発明に係る定常雑音下における音声区間検出に基づく目的音声の復元方法において、前記検出判定基準は、前記総和 $F$ が前記閾値 $\beta$ を超える周波数領域を前記音声区間とし、前記総和 $F$ が前記閾値 $\beta$ 以下となる周波数領域を前記雑音区間として検出するように設定されていることが好ましい。

これによって、検出判定基準として2値化された（閾値 $\beta$ に基づいて音声区間とするか、雑音区間とするかの2者択一の選択を行う）音声区間検出閾数を設定することができる。その結果、各推定スペクトル $Y^*$ から音声の特徴付けるスペクトルの現れる周波数領域を簡便に決定することができる。

#### 【0010】

前記目的に沿う第2の発明に係る定常雑音下における音声区間検出に基づく目的音声の復元方法は、異なる2つの音源からそれぞれ発信される目的音声及び雑音を、異なる位置に設けた第1及び第2のマイクでそれぞれ受信して混合信号を形成し、該各混合信号を時間領域から周波数領域にフーリエ変換して、独立成分分析法により前記目的音声及び前記雑音にそれぞれ対応する各分割スペクトルから複数の推定スペクトル $Y^*$ 、 $Y$ を抽出する第1工程と、

前記各推定スペクトル $Y^*$ を、該各推定スペクトル $Y^*$ の振幅分布の尖度の値を分離判定基準にして前記雑音が除去された複数の推定スペクトル $y^*$ 及び前記雑音が残留する複数の推定スペクトル $y$ に分離する第2工程と、

前記各推定スペクトル $y^*$ の総和 $F$ を求め、該総和 $F$ の最大値に応じて決定される閾値 $\beta$ を検出判定基準にして該総和 $F$ の時間領域から音声区間及び雑音区間をそれぞれ検出する第3工程と、

前記各推定スペクトル $Y^*$ を周波数領域から時間領域にフーリエ逆変換して得られた前記目的音声の信号から前記音声区間に存在する成分を抽出して前記目的音声を復元する第4工程とを有する。

#### 【0011】

混合信号を時間領域から周波数領域にフーリエ変換すると、混合信号は離散化された複数のスペクトルに分解される。そして、各スペクトルを大きさの順に並べたときのフレーム番号とサンプリング時刻の間には、フレーム周期を介して1対1の関係が成立する。従って、フレーム番号とサンプリング時刻の関係から、周波数領域で検出した音声区間を時間領域に変換することができる。そして、音声区間を除いた部分を時間領域における雑音区

間として検出することができる。

このため、各推定スペクトル  $Y^*$  を周波数領域から時間領域にフーリエ逆変換して得られた時間領域の目的音声の信号から音声区間に存在する成分のみを抽出して目的音声の復元信号を生成することができる。

#### 【0012】

第2の発明に係る定常雑音下における音声区間検出に基づく目的音声の復元方法において、前記検出判定基準は、前記総和  $F$  が前記閾値  $\beta$  を超える時間領域を前記音声区間とし、前記総和  $F$  が前記閾値  $\beta$  以下となる時間領域を前記雑音区間として検出することが好ましい。

これによって、検出判定基準として2値化された（閾値  $\beta$  に基づいて音声区間とするか、雑音区間とするかの2者択一の選択を行う）音声区間検出閾数を設定することができる。その結果、時間領域の目的音声の信号から雑音区間がゼロになって音声区間のみの音声強調された信号を簡便に決定することができる。

#### 【0013】

第1及び第2の発明に係る定常雑音下における音声区間検出に基づく目的音声の復元方法において、前記尖度の値を前記各推定スペクトル  $Y^*$  の振幅分布のエントロピー  $H$  で評価することが好ましい。

各推定スペクトル  $Y^*$  の振幅分布の尖度は、振幅分布の不確定性の観点からエントロピー  $H$  を用いて定量的に評価することができる。この場合、雑音が除去される程推定スペクトル  $Y^*$  のエントロピー  $H$  は小さくなり、雑音が残留する程推定スペクトル  $Y^*$  のエントロピー  $H$  は大きくなる。

なお、尖度を定量的に計る尺度としては、例えば、平均の周りの四次のモーメント  $\mu$  と標準偏差  $\sigma$  の4乗との比、すなわち  $\mu/\sigma^4$  (kurtosis) を使用することもできるが、異常値に対する評価の安定性が劣るという問題があり好ましくない。

#### 【0014】

第1及び第2の発明に係る定常雑音下における音声区間検出に基づく目的音声の復元方法において、前記分離判定基準は、

(1) 前記エントロピー  $H$  が予め設定した閾値  $\alpha$  未満である推定スペクトル  $Y^*$  は前記推定スペクトル  $y^*$  として分離し、

(2) 前記エントロピー  $H$  が予め設定した閾値  $\alpha$  以上である推定スペクトル  $Y^*$  は前記推定スペクトル  $y$  として分離するように設定されていることが好ましい。

#### 【0015】

各推定スペクトル  $Y^*$  を周波数毎に見てみると、雑音が良好に分離されている周波数と雑音がほとんど除去されていない周波数が存在しているので、推定スペクトル  $Y^*$  のエントロピー  $H$  は分布を持つことになる。

従って、エントロピー  $H$  の値に閾値  $\alpha$  を設定して、エントロピー  $H$  がこの閾値  $\alpha$  未満である推定スペクトル  $Y^*$  は雑音の除去が行われている推定スペクトル  $y^*$  と認定し、エントロピー  $H$  がこの閾値  $\alpha$  以上である推定スペクトル  $Y^*$  は雑音が残留する推定スペクトル  $y$  と認定することができる。

このように、2値化された（閾値  $\alpha$  に基づいて推定スペクトル  $y^*$  とするか、推定スペクトル  $y$  とするか）2者択一の選択を行う）分離判定基準を設定することにより、各推定スペクトル  $Y^*$  から推定スペクトル  $y^*$  及び推定スペクトル  $y$  を素早く、かつ容易に決定することができる。

#### 【発明の効果】

#### 【0016】

請求項1、及びこれに従属する2、5、6記載の定常雑音下における音声区間検出に基づく目的音声の復元方法においては、実際的な環境下で受信した信号から目的音声に対応する分割スペクトルを抽出し音声区間と雑音区間を検出して音声区間に含まれる信号のみを抽出することができるので、音声全体に含まれる残留雑音を最小化して高品質の音声を復元することが可能になる。



その結果、雑音環境下での音声認識として、例えば、OA分野での音声指令、音声入力、流通業界での倉庫管理やカーナビゲーターへの音声指令等、従来のタッチセンサ、指、及びキーボードを用いた入力操作の代替が可能になる。

#### 【0017】

特に、請求項2記載の定常雑音下における音声区間検出に基づく目的音声の復元方法においては、各推定スペクトル $Y^*$ から音声の特徴付けるスペクトルの現れる周波数領域を簡便に決定することができるので、素早く音声区間を検出することが可能になる。その結果、実環境下で音声復元の応答速度が速く、しかも、高品質の音声認識エンジンを供給することが可能になる。

#### 【0018】

請求項3、及びこれに従属する4～6記載の定常雑音下における音声区間検出に基づく目的音声の復元方法においては、実際の環境下で受信した信号から抽出した目的音声の時間信号に対して音声区間と雑音区間を検出して音声区間に含まれる信号のみを抽出することができるので、音声全体に含まれる残留雑音を最小化して高品質の音声を復元することが可能になる。

その結果、雑音環境下での音声認識として、例えば、OA分野での音声指令、音声入力、流通業界での倉庫管理やカーナビゲーターへの音声指令等、従来のタッチセンサ、指、及びキーボードを用いた入力操作の代替が可能になる。

#### 【0019】

特に、請求項4記載の定常雑音下における音声区間検出に基づく目的音声の復元方法においては、目的音声の時間信号から音声の特徴付ける信号の現れる時間領域を少ない計算量で検出することができるので、素早く音声区間を検出することが可能になる。その結果、実環境下で音声復元の応答速度が速く、しかも、高品質の音声認識エンジンを供給することが可能になる。

#### 【0020】

請求項5記載の定常雑音下における音声区間検出に基づく目的音声の復元方法においては、各推定スペクトル $Y^*$ に異常値が含まれていても尖度を確実に評価することができるので、雑音が除去された推定スペクトル $y^*$ 及び雑音が残留する推定スペクトル $y$ をそれぞれ確実に分離することが可能になる。

#### 【0021】

請求項6記載の定常雑音下における音声区間検出に基づく目的音声の復元方法においては、推定スペクトル $Y^*$ から推定スペクトル $y^*$ 及び推定スペクトル $y$ を少ない計算量で精度良く抽出することができるので、実環境下で音声復元の応答速度が速く、しかも、認識能力の極めて高い音声認識エンジンを供給することが可能になる。

#### 【発明を実施するための最良の形態】

#### 【0022】

続いて、添付した図面を参照しつつ、本発明を具体化した実施の形態につき説明し、本発明の理解に供する。

図1は本発明の第1及び第2の実施の形態に係る定常雑音下における音声区間検出に基づく目的音声の復元方法を適用した目的音声の復元装置の構成図、図2は同復元方法における目的音声及び雑音から各推定スペクトル $Y^*$ が形成されるまでの信号の流れを示す説明図、図3各推定スペクトル $Y^*$ から構成された復元スペクトル群を周波数領域から時間領域にフーリエ逆変換して得られた目的音声信号の波形を示す説明図、図4は雑音が除去された推定スペクトル $Y^*$ の波形を示す説明図、図5は雑音が残留している推定スペクトル $Y^*$ の波形を示す説明図、図6は雑音が除去された推定スペクトル $Y^*$ の振幅分布を示す説明図、図7は雑音が残留している推定スペクトル $Y^*$ の振幅分布を示す説明図、図8は雑音が除去された各推定スペクトル $y^*$ の総和を示す説明図、図9は各推定スペクトル $Y^*$ に対して決定された音声区間検出関数のグラフ、図10は各推定スペクトル $Y^*$ から音声区間に存在する成分を抽出して生成した目的音声の復元スペクトル群を周波数領域から時間領域にフーリエ逆変換して得られた目的音声信号の説明図、図11は本発明の実施例

1、2におけるバーチャルルームでの音源とマイクの配置を示す説明図である。

【0023】

図1に示すように、本発明の第1及び第2の実施の形態に係る定常雑音下における音声区間検出に基づく目的音声の復元方法を適用した目的音声の復元装置10は、異なる2つの音源11、12（一方が目的音声音源で他方が雑音源であるが、特定されていない）からそれぞれ発信される信号を受信する異なる位置に設けた第1のマイク13及び第2のマイク14と、各マイク13、14で受信されて得られた混合信号をそれぞれ増幅する第1及び第2の増幅器15、16と、各増幅器15、16から入力される混合信号から目的音声及び雑音を分離して復元信号として出力する復元装置本体17を有している。更に、目的音声の復元装置10は、出力された復元信号を増幅する復元信号増幅器18と、増幅された復元信号を出力するスピーカ19を有している。以下、これらについて詳細に説明する。

【0024】

第1及び第2のマイク13、14としては、例えば、可聴音域（10～20000Hz）の信号を集音するのに十分な周波数特性を有するマイクを使用することができる。ここで、第1のマイク13は、音源11に対して、第2のマイク14よりも接近して配置され、第2のマイク14は、音源12に対して、第1のマイク13よりも接近して配置されている。

また、第1及び第2の増幅器15、16としては、可聴音域の信号を歪みなく増幅可能な周波数帯域の特性を備えた増幅器を使用できる。

【0025】

復元装置本体17は、各増幅器15、16から入力される混合信号をデジタル化するためのA/D変換器20、21を有している。

また、復元装置本体17は、デジタル化された各混合信号を時間領域から周波数領域にフーリエ変換し、独立成分分析法の一例であるFastICA法により2つの分離信号 $U_1$ 、 $U_2$ に分解する分離信号作成演算回路と、音源11及び音源12から第1及び第2のマイク13、14までの各伝達経路特性に基づいて、分離信号 $U_1$ から第1のマイク13で受信された複数のスペクトル $v_{11}$ と第2のマイク14で受信された複数のスペクトル $v_{12}$ を生成し、分離信号 $U_2$ から第1のマイク13で受信された複数のスペクトル $v_{21}$ と第2のマイク14で受信された複数のスペクトル $v_{22}$ を生成して分割スペクトルを形成する分割スペクトル生成演算回路を備えた分割スペクトル生成器22を有している。

【0026】

更に、復元装置本体17は、分割スペクトル生成器22で生成させた各スペクトル $v_{11}$ 、 $v_{12}$ 、 $v_{21}$ 、 $v_{22}$ に対して、第1及び第2のマイク13、14と音源11及び音源12との間の距離に基づく音の伝達特性を用いた判定基準を適用して、各分割スペクトルを目的音声及び雑音にそれぞれ対応させ、目的音声を復元するための複数の推定スペクトル $Y^*$ を抽出し出力する推定スペクトル抽出回路23と、推定スペクトル抽出回路23から出力された各推定スペクトル $Y^*$ の振幅分布の尖度の値を分離判定基準にして、各推定スペクトル $Y^*$ を雑音が除去された複数の推定スペクトル $y^*$ と雑音が残留する複数の推定スペクトル $y$ に分離し、各推定スペクトル $y^*$ の総和 $F$ を求め総和 $F$ の最大値に応じて決定される閾値 $\beta$ を検出判定基準にして、総和 $F$ の周波数領域から音声区間を検出し出力する音声区間検出回路24を有している。また、復元装置本体17は、各推定スペクトル $Y^*$ から検出された音声区間に存在する成分を抽出して目的音声の復元スペクトル群を生成して出力する復元スペクトル抽出回路25を有している。

そして、更に、復元装置本体17は、復元スペクトル抽出回路25から出力された復元スペクトル群を周波数領域から時間領域にフーリエ逆変換して復元信号を生成させる復元信号生成回路26を有している。

【0027】

そして、分離信号作成演算回路及び分割スペクトル生成演算回路を備えた分割スペクトル生成器22、推定スペクトル抽出回路23、音声区間検出回路24、復元スペクトル抽出

回路 25、及び復元信号生成回路 26 は、各回路の機能を発現する各プログラムを、例えば、パーソナルコンピュータに搭載させることにより構成することができる。また、各プログラムをマイクロコンピュータに搭載させ、これらのマイクロコンピュータが連携動作可能なように回路を形成することにより構成することもできる。

特に、パーソナルコンピュータに各プログラムを搭載させた場合は、このパーソナルコンピュータに A/D 変換器 20、21 を取付けることにより、復元装置本体 17 を一括して構成することができる。

また、復元信号増幅器 18 は、復元信号をアナログ変換して可聴音域を歪みなく増幅することが可能な特性を備えた増幅器を使用することができ、スピーカ 19 も可聴音域の信号を歪みなく出力可能な特性を備えたスピーカを使用できる。

#### 【0028】

次に、本発明の第 1 の実施の形態に係る定常雑音下における音声区間検出に基づく目的音声の復元方法は、図 1 に示すように、異なる 2 つの音源 11、12 からそれぞれ発信される信号  $s_1(t)$  及び信号  $s_2(t)$  (一方が目的音声信号であり、他方が雑音信号である) を、異なる位置に設けた第 1 及び第 2 のマイク 13、14 でそれぞれ受信して混合信号  $x_1(t)$ 、 $x_2(t)$  を形成し、各混合信号  $x_1(t)$ 、 $x_2(t)$  を時間領域から周波数領域にフーリエ変換して、Fast ICA 法により目的音声及び雑音にそれぞれ対応する複数の推定スペクトル  $Y^*$ 、 $Y$  を抽出する第 1 工程を有している。

また、目的音声の復元方法は、各推定スペクトル  $Y^*$  を、各推定スペクトル  $Y^*$  の振幅分布の尖度の値を分離判定基準にして雑音が除去された複数の推定スペクトル  $y^*$  及び雑音が残留する複数の推定スペクトル  $y$  に分離する第 2 工程と、各推定スペクトル  $y^*$  の総和  $F$  を求め、総和  $F$  の最大値に応じて決定される閾値  $\beta$  を検出判定基準にして総和  $F$  の周波数領域から音声区間及び雑音区間をそれぞれ検出する第 3 工程を有している。

更に、目的音声の復元方法は、各推定スペクトル  $Y^*$  から検出した音声区間に存在する成分を抽出して目的音声の復元スペクトル群を生成し、復元スペクトル群を周波数領域から時間領域にフーリエ逆変換して目的音声を復元する第 4 工程とを有する。なお、 $t$  は時間を示す。以下、これらの各工程について詳細に説明する。

#### 【0029】

##### (第 1 工程)

音源 11 から発信される信号  $s_1(t)$  と、音源 12 から発信される雑音信号  $s_2(t)$  は、一般に統計的に独立と考えることができる。そして、信号  $s_1(t)$  と信号  $s_2(t)$  を、異なる位置に設置した第 1 及び第 2 のマイク 13、14 で受信して得られる各混合信号  $x_1(t)$ 、 $x_2(t)$  は、式 (1) のように表記できる。

ここで、 $s(t) = [s_1(t), s_2(t)]^T$ 、 $x(t) = [x_1(t), x_2(t)]^T$ 、 $*$  は畳み込み記号、 $G(t)$  は各音源 11、12 から各マイク 13、14 までのそれぞれの伝達関数である。

#### 【0030】

##### 【数 1】

$$x(t) = G(t) * s(t) \quad \dots\dots (1)$$

#### 【0031】

(1) 式のように、各音源 11、12 からの信号が畳み込まれて観測される場合、各混合信号  $x_1(t)$ 、 $x_2(t)$  から、信号  $s_1(t)$  と信号  $s_2(t)$  を、時間領域で分離することは困難となる。そのため、混合信号  $x_1(t)$ 、 $x_2(t)$  を式 (2) のように短時間間隔 (フレーム)、例えば数 10 msec 程度の時間間隔で分割し、各フレーム毎に時間領域から周波数領域にフーリエ変換して扱う。周波数領域にフーリエ変換することにより、瞬時混合のときの問題と同様に扱える。

また、得られた各周波数におけるスペクトルをフレーム順に並べることにより、スペクトルを時系列として扱うことができる。

#### 【0032】

## 【数2】

$$x_j(\omega, k) = \sum_t e^{-j\omega t} x_j(t) w(t - k\tau) \quad \dots\dots (2)$$

$$(j=1, 2; k=0, 1, \dots, K-1)$$

## 【0033】

ここに、 $\omega (=0, 2\pi/M, \dots, 2\pi(M-1)/M)$  は規格化周波数、 $M$  はフレーム内のサンプル数、 $w(t)$  は窓関数、 $\tau$  はフレーム周期、 $K$  はフレーム数を表す。このとき混合信号スペクトル  $x(\omega, k)$  と、信号  $s_1(t)$  及び信号  $s_2(t)$  の各スペクトルは、周波数領域で式(3)のように関係づけられる。ここに、 $s(\omega, k)$  は  $s(t)$  を窓掛けして離散フーリエ変換したもので、 $G(\omega)$  は  $G(t)$  を離散してフーリエ変換して得られる複素定数行列である。

## 【0034】

ここで、信号スペクトル  $s_1(\omega, k)$  及び信号スペクトル  $s_2(\omega, k)$  は本来独立であるので、Fast ICA法を用いて、混合信号スペクトル  $x(\omega, k)$  から、互いに独立となる分離信号スペクトル  $U_1(\omega, k)$ 、 $U_2(\omega, k)$  を求めると、これらのスペクトルが信号スペクトル  $s_1(\omega, k)$  及び信号スペクトル  $s_2(\omega, k)$  に相当することになる。

すなわち、混合信号スペクトル  $x(\omega, k)$  と分離信号スペクトル  $U_1(\omega, k)$ 、 $U_2(\omega, k)$  の間に、式(4)の関係が成立するような分離行列  $H(\omega)$  を求めることにより、混合信号スペクトル  $x(\omega, k)$  から、互いに独立となる分離信号スペクトル  $U_1(\omega, k)$ 、 $U_2(\omega, k)$  を決定することができる。ここで、 $u(\omega, k) = [U_1(\omega, k), U_2(\omega, k)]^T$  である。

## 【0035】

## 【数3】

$$x(\omega, k) = G(\omega) s(\omega, k) \quad \dots\dots (3)$$

## 【0036】

## 【数4】

$$u(\omega, k) = H(\omega) x(\omega, k) \quad \dots\dots (4)$$

## 【0037】

なお、周波数領域では、個々の周波数  $\omega$  で、式(5)のように振幅の曖昧さや成分置換の問題が生じる。そのため、復元するのに意味ある分離信号を得るには、これらの問題を解決する必要がある。

ここに  $Q(\omega)$  は白色化行列、 $P$  は各行と列の全ての要素が値1である1つの要素を除いて0である成分置換を表す行列、 $D(\omega) = \text{diag}[d_1(\omega), d_2(\omega)]$  は振幅の曖昧さを表す対角行列である。

## 【0038】

## 【数5】

$$H(\omega) Q(\omega) G(\omega) = PD(\omega) \quad \dots\dots (5)$$

## 【0039】

次に、周波数領域で、各信号スペクトル  $s_i(\omega, k)$  ( $i=1, 2$ ) を、その実部と虚部は平均がゼロで等しい分散をもち、実部と虚部は無相関という仮定の下で次のように定式化する。すなわち、周波数  $\omega$  において、分離荷重  $h_n(\omega)$  ( $n=1, 2$ ) を式(6)、式(7)に示すFast ICA法のアルゴリズムに従って更新する。

ここに、 $f(\cdot)$  は非線型関数で、 $f'(\cdot)$  は  $f(\cdot)$  の微分、 $\bar{\cdot}$  は共役、 $K$  はフレームのサンプル数である。

【0040】

【数6】

$$h_n^+(\omega) = \frac{1}{K} \sum_{k=0}^{K-1} \{ x(\omega, k) \bar{u}_n(\omega, k) f(|u_n(\omega, k)|^2) - [f(|u_n(\omega, k)|^2) + |u_n(\omega, k)|^2 f'(|u_n(\omega, k)|^2)] h_n(\omega) \} \dots\dots (6)$$

【0041】

【数7】

$$h_n(\omega) = h_n^+(\omega) / \| h_n^+(\omega) \| \dots\dots (7)$$

【0042】

このアルゴリズムは、式(8)に示す収束条件CCがほぼ1(例えば、CCが0.9999以上)を満たすまで、繰り返される。さらに、 $h_2(\omega)$ については、式(9)のように、 $h_1(\omega)$ と直交化させて再び(7)式により規格化する。

【0043】

【数 8】

$$CC = \bar{h}_n^T(\omega) h_n^+(\omega) \simeq 1 \quad \dots\dots (8)$$

【0044】

【数 9】

$$h_2(\omega) = h_2(\omega) - h_1(\omega) \bar{h}_1^T(\omega) h_2(\omega) \quad \dots\dots (9)$$

【0045】

上述の Fast ICA 法のアルゴリズムを各周波数  $\omega$  について適用し、得られる分離荷重  $h_n(\omega)$  ( $n=1, 2$ ) を式 (4) の  $H(\omega)$  に対して、式 (10) として代入すれば、各周波数での分離信号スペクトル  $u(\omega, k) = [U_1(\omega, k), U_2(\omega, k)]^T$  が求まる。

【0046】

【数 10】

$$H(\omega) = [\bar{h}_1^T(\omega), \bar{h}_2^T(\omega)]^T \quad \dots\dots (10)$$

【0047】

図 2 に示すように、分離信号スペクトル  $U_1(\omega, k)$ 、 $U_2(\omega, k)$  が出力される 2 つのノードを 1、2 と表記する。

このとき、分割スペクトル  $v_1(\omega, k) = [v_{11}(\omega, k), v_{12}(\omega, k)]^T$ 、 $v_2(\omega, k) = [v_{21}(\omega, k), v_{22}(\omega, k)]^T$  は、分離信号スペクトル  $U_n(\omega, k)$  から、式 (11)、(12) に示すように、各ノード  $n (=1, 2)$  で対になって生成されるスペクトルとして定義する。

【0048】

【数 11】

$$\begin{bmatrix} v_{11}(\omega, k) \\ v_{12}(\omega, k) \end{bmatrix} = (H(\omega) Q(\omega))^{-1} \begin{bmatrix} U_1(\omega, k) \\ 0 \end{bmatrix} \quad \dots\dots (11)$$

【0049】

【数 12】

$$\begin{bmatrix} v_{21}(\omega, k) \\ v_{22}(\omega, k) \end{bmatrix} = (H(\omega) Q(\omega))^{-1} \begin{bmatrix} 0 \\ U_2(\omega, k) \end{bmatrix} \quad \dots\dots (12)$$

【0050】

ここで、成分置換は生じていないが、振幅の曖昧さが存在する場合、分離信号スペクトル  $U_n(\omega, k)$  は、式 (13) として出力される。そして、この分離信号  $U_n(\omega, k)$  に対する分割スペクトルは、信号スペクトル  $s_1(\omega, k)$  及信号スペクトル  $s_2(\omega, k)$  と伝達関数との積として、式 (14)、式 (15) のように生成される。

なお、 $g_{11}(\omega)$  は音源 11 から第 1 のマイク 13 までの伝達関数、 $g_{21}(\omega)$  は音源 11 から第 2 のマイク 14 までの伝達関数、 $g_{12}(\omega)$  は音源 12 から第 1 のマイク 13 までの伝達関数、 $g_{22}(\omega)$  は音源 12 から第 2 のマイク 14 までの伝達関数を示す。

【0051】

【数 1 3】

$$\begin{bmatrix} U_1(\omega, k) \\ U_2(\omega, k) \end{bmatrix} = \begin{bmatrix} d_1(\omega) s_1(\omega, k) \\ d_2(\omega) s_2(\omega, k) \end{bmatrix} \dots\dots (13)$$

【0052】

【数 1 4】

$$\begin{bmatrix} v_{11}(\omega, k) \\ v_{12}(\omega, k) \end{bmatrix} = \begin{bmatrix} g_{11}(\omega) s_1(\omega, k) \\ g_{21}(\omega) s_1(\omega, k) \end{bmatrix} \dots\dots (14)$$

【0053】

【数 1 5】

$$\begin{bmatrix} v_{21}(\omega, k) \\ v_{22}(\omega, k) \end{bmatrix} = \begin{bmatrix} g_{12}(\omega) s_2(\omega, k) \\ g_{22}(\omega) s_2(\omega, k) \end{bmatrix} \dots\dots (15)$$

【0054】

また、成分置換と振幅の曖昧さの両方がある場合、分離信号スペクトル  $U_n(\omega, k)$  は、式(16)となつて、ノード1、2での分割スペクトルは、式(17)、式(18)のように生成される。

なお、ノード1で生成するスペクトル  $v_{11}(\omega, k)$  は、音源12から発信された信号スペクトル  $s_2(\omega, k)$  を第1のマイク13で観測したときのスペクトル、ノード1で生成するスペクトル  $v_{12}(\omega, k)$  は、音源12から発信された信号スペクトル  $s_2(\omega, k)$  を第2のマイク14で観測したときのスペクトルを示す。また、ノード2で生成するスペクトル  $v_{21}(\omega, k)$  は、音源11から発信された信号スペクトル  $s_1(\omega, k)$  を第1のマイク13で観測したときのスペクトル、ノード2で生成するスペクトル  $v_{22}(\omega, k)$  は、音源11から発信された信号スペクトル  $s_1(\omega, k)$  を第2のマイク14で観測したときのスペクトルを示す。

【0055】

【数 1 6】

$$\begin{bmatrix} U_1(\omega, k) \\ U_2(\omega, k) \end{bmatrix} = \begin{bmatrix} d_1(\omega) s_2(\omega, k) \\ d_2(\omega) s_1(\omega, k) \end{bmatrix} \dots\dots (16)$$

【0056】

【数 1 7】

$$\begin{bmatrix} v_{11}(\omega, k) \\ v_{12}(\omega, k) \end{bmatrix} = \begin{bmatrix} g_{12}(\omega) s_2(\omega, k) \\ g_{22}(\omega) s_2(\omega, k) \end{bmatrix} \dots\dots (17)$$

【0057】

【数 1 8】

$$\begin{bmatrix} v_{21}(\omega, k) \\ v_{22}(\omega, k) \end{bmatrix} = \begin{bmatrix} g_{11}(\omega) s_1(\omega, k) \\ g_{21}(\omega) s_1(\omega, k) \end{bmatrix} \dots\dots (18)$$

【0058】

図2に示す4つのスペクトル  $v_{11}(\omega, k)$ 、 $v_{12}(\omega, k)$ 、 $v_{21}(\omega, k)$ 、 $v_{22}(\omega, k)$  は、1つの音源に対して2つの候補となる推定値を有している。そして、一方の

音源からの信号に対して成分置換がない場合には  $v_{11}(\omega, k)$  と  $v_{12}(\omega, k)$  があり、成分置換がある場合には  $v_{21}(\omega, k)$  と  $v_{22}(\omega, k)$  がある。

ここで、第1及び第2のマイクと各音源との距離に基づく音の伝達特性、例えば、音の強弱を考慮すると、周波数毎に得られた各スペクトル  $v_{11}$ 、 $v_{12}$ 、 $v_{21}$ 、 $v_{22}$  間に、スペクトル強度の差が生じる。従って、第1及び第2のマイク13、14と各音源との距離に明確な差を予め設けておくと、各音源がどのマイクで受信されたか、すなわち、2つの音源の信号がそれぞれどのスペクトル  $v_{11}$ 、 $v_{12}$ 、 $v_{21}$ 、 $v_{22}$  に相当するかを判定することができる。

例えば、音源11は第2のマイク14に比べて第1のマイク13に近く、音源12は第1のマイク13に比べて第2のマイク14に近いと仮定する。このとき、音源11から各マイク13、14への信号の伝達特性を比較すると、式(19)なるゲインの大小関係が得られる。同様に、音源12から各マイク13、14への伝達特性についても、式(20)なるゲインの大小関係が得られる。

【0059】

【数19】

$$|g_{11}(\omega)| > |g_{21}(\omega)| \quad \dots\dots (19)$$

【0060】

【数20】

$$|g_{12}(\omega)| < |g_{22}(\omega)| \quad \dots\dots (20)$$

【0061】

このとき、式(19)、(20)のゲインの大小関係を念頭に、式(14)、(15)、もしくは式(17)、(18)を用いて、スペクトル  $v_{11}$  とスペクトル  $v_{12}$  との差  $D_1$ 、及びスペクトル  $v_{21}$  とスペクトル  $v_{22}$  との差  $D_2$  を求めると、成分置換がない場合、ノード1での差  $D_1$  は正、またノード2で差  $D_2$  は負となることが導かれる。

一方、成分置換がある場合、同様にしてスペクトル  $v_{11}$  とスペクトル  $v_{12}$  との差  $D_1$ 、及びスペクトル  $v_{21}$  とスペクトル  $v_{22}$  との差  $D_2$  を考えると、ノード1での差  $D_1$  は負、またノード2での差  $D_2$  は正となることが導かれる。

従って、成分置換の有無については、分割スペクトルの差をそれぞれ調べて、ノード1での差  $D_1$  が正かつノード2での差  $D_2$  が負のとき、成分置換はないと判定し、ノード1での差  $D_1$  が負かつノード2での差  $D_2$  が正のとき、成分置換はあると判定する。

【0062】

ここで、差  $D_1$  をスペクトル  $v_{11}$  とスペクトル  $v_{21}$  との各絶対値の差として、差  $D_2$  をスペクトル  $v_{21}$  とスペクトル  $v_{22}$  との各絶対値の差としてそれぞれ求めた場合、差  $D_1$ 、差  $D_2$  はそれぞれ、式(21)、式(22)のように表される。

【0063】

【数21】

$$D_1 = |v_{11}(\omega, k)| - |v_{12}(\omega, k)| \quad \dots\dots (21)$$

【0064】

【数22】

$$D_2 = |v_{21}(\omega, k)| - |v_{22}(\omega, k)| \quad \dots\dots (22)$$

【0065】

ここで、成分置換がない場合には、一方の音源の信号のスペクトル  $y_1(\omega, k)$  に、 $v_{11}(\omega, k)$  を採用する。これは、 $v_{11}(\omega, k)$  は第1のマイク13で観測されたスペクトルの推定値であり、第2のマイク14で観測されたスペクトルの推定値  $v_{12}(\omega, k)$



より大きく、周囲の暗騒音の影響を受けにくいからである。また、成分置換がある場合には、一方の音源のスペクトル  $y_1(\omega, k)$  に  $v_{21}(\omega, k)$  を採用する。

以上のことから、一方の音源のスペクトル  $y_1(\omega, k)$  は式(23)として表せる。同様に、他方の音源のスペクトル  $y_2(\omega, k)$  に、成分置換がない場合には  $v_{22}(\omega, k)$  を採用し、成分置換がある場合には  $v_{12}(\omega, k)$  を採用する。以上のことから、他方の音源のスペクトル  $y_2(\omega, k)$  は式(24)として表せる。

なお、成分置換の有無は、式(21)、(22)を使用し判定する。

【0066】

【数23】

$$y_1(\omega, k) = \begin{cases} v_{11}(\omega, k) & \text{if } D_1 > 0, D_2 < 0 \\ v_{21}(\omega, k) & \text{if } D_1 < 0, D_2 > 0 \end{cases} \dots\dots (23)$$

【0067】

【数24】

$$y_2(\omega, k) = \begin{cases} v_{12}(\omega, k) & \text{if } D_1 < 0, D_2 > 0 \\ v_{22}(\omega, k) & \text{if } D_1 > 0, D_2 < 0 \end{cases} \dots\dots (24)$$

【0068】

FastICA法は、混合信号から非ガウス性の高い順に音源を分離するという特徴を有している。また、一般に、話者音声は雑音より非ガウス性が高いという特徴を有している。これらのことは、音源が話者音声と騒音である場合、話者音声に対応する分割スペクトルは、最初に出力される分離信号  $U_1$  に出力される頻度が高くなる。

従って、一方の音源が話者音声であれば、成分置換のない頻度は高く、逆に、他方の音源が話者音声であれば、成分置換の生じる頻度は高くなる。

【0069】

このことから、フーリエ変換した際の各周波数毎に各スペクトル  $y_1$ 、 $y_2$  を生成するとき、成分置換の有無の頻度、すなわち個数  $N_+$ 、 $N_-$  を求め、個数  $N_+$  が個数  $N_-$  より大きい場合、推定スペクトル  $Y^*$  としてスペクトル  $y_1$  を抽出し、推定スペクトル  $Y$  としてスペクトル  $y_2$  を抽出する。一方、個数  $N_-$  が個数  $N_+$  より大きい場合、推定スペクトル  $Y^*$  としてスペクトル  $y_2$  を抽出し、推定スペクトル  $Y$  としてスペクトル  $y_1$  を抽出する。

【0070】

(第2工程)

以上のようにして周波数毎に抽出した各推定スペクトル  $Y^*$  から構成される復元スペクトル群を周波数領域から時間領域にフーリエ逆変換して得られた目的音声信号の波形(東京という音声)を図3に示す。図3から、復元された目的音声信号には、雑音信号が残留していることが判る。

そこで、抽出された複数の推定スペクトル  $Y^*$  を周波数毎に見てみると、図4に示すように、雑音が除去された推定スペクトル  $Y^*$  と、図5に示すように、雑音が残留している推定スペクトル  $Y^*$  とが存在している。そして、図4に示される雑音が除去された推定スペクトル  $Y^*$  の波形では、音声区間では振幅が大きく、非音声区間では振幅が非常に小さくなって、音声の始端と終端が明確に検出できる。従って、雑音が除去された推定スペクトル  $Y^*$  のみを利用すると精度の高い音声区間の検出できると考えられる。

【0071】

ここで、図4、図5に示す各推定スペクトル  $Y^*$  の振幅分布をそれぞれ求めると、図6、図7に示すようになる。これらの図から、雑音が除去された推定スペクトル  $Y^*$  の振幅分布の尖度が高く、雑音が残留している推定スペクトル  $Y^*$  の振幅分布の尖度は低いことが判る。

従って、推定スペクトル  $Y^*$  の振幅分布の尖度を分離判定基準にすると、尖度の値の違いにより、各推定スペクトル  $Y^*$  を雑音が除去されている複数の推定スペクトル  $y^*$  と雑音が除去されていない複数の推定スペクトル  $y$  とにそれぞれ分離できる。

【0072】

尖度の値を定量的に評価する方法として、振幅分布のエントロピー  $H$  を使用することができる。エントロピー  $H$  は振幅分布の不確実性を示す尺度であり、尖度が高ければエントロピーは小さく、尖度は低ければエントロピーは大きくなる。従って、エントロピー  $H$  の値に予め閾値  $\alpha$  を設定しておく、分離判定基準は、(1) エントロピー  $H$  が閾値  $\alpha$  未満である推定スペクトル  $Y^*$  は推定スペクトル  $y^*$  として分離し、(2) エントロピー  $H$  が閾値  $\alpha$  以上である推定スペクトル  $Y^*$  は推定スペクトル  $y$  として分離することになる。

なお、エントロピー  $H$  は式 (25) で算出する。

【0073】

【数25】

$$H(\omega) = - \sum_{n=1}^N p_{\omega}(l_n) \log p_{\omega}(l_n) \quad \dots\dots (25)$$

【0074】

ここに、 $p_{\omega}(l_n)$  は、各推定スペクトル  $Y^*$  の実部の値の分布範囲を  $N$  等分したとき区画  $l_n$  に入る頻度  $q_{\omega}(l_n)$  ( $n=1, 2, \dots, N$ ) を、式 (26) のように規格化して得られる確率である。

【0075】

【数26】

$$p_{\omega}(l_n) = q_{\omega}(l_n) / \sum_{n=1}^N q_{\omega}(l_n) \quad \dots\dots (26)$$

【0076】

(第3工程)

音声の周波数成分は、時間的に変化するため、音声の特徴付けるスペクトルの現れる周波数領域は推定スペクトル  $y^*$  によって異なる。そこで、各推定スペクトル  $y^*$  の総和  $F$  を求めることにより、音声の特徴付けるスペクトルの現れる周波数領域の範囲が明確になり、時間的な音声区間と一致する周波数領域における音声区間を検出することができる。

算出される総和  $F$  は、例えば図8に示すような分布をもつ。なお、図8では、総和  $F$  の最大値で規格化して表示している。そこで、総和  $F$  の最大値 (図8では1) に応じて決定される閾値  $\beta$  を適切に選ぶことにより、総和  $F$  が閾値  $\beta$  を超える領域を音声区間、総和  $F$  が閾値  $\beta$  以下の領域を雑音区間として特定することができる。

従って、図8の振幅分布と閾値  $\beta$  を用いて、検出判定基準として、図9に示すような2者択一型の音声区間検出関数  $F^*(k)$  を設定することができる。ここで、 $F^*(k)$  は、閾値  $\beta < F$  の領域で  $F^*(k) = 1$ 、閾値  $\beta \geq F$  の領域で  $F^*(k) = 0$  となる2値化関数である。

【0077】

(第4工程)

従って、目的音声に対応する各推定スペクトル  $Y^*$  に対して音声区間検出関数を掛けることにより、各推定スペクトル  $Y^*$  から音声区間に存在する成分のみを抽出することができる。

その結果、各推定スペクトル  $Y^*$  の音声区間に存在する成分のみから目的音声の復元スペクトル群  $\{Z(\omega, k) \mid k=0, 1, \dots, K-1\}$  を生成して、フレーム毎に逆離散フーリエ変換 (フーリエ逆変換) して時間領域に戻して、それをすべてのフレームにわたって式 (27) のように総和を取れば、目的音声音源の復元信号  $Z(t)$  が得られる。

図10に、図3の目的音声信号の復元に使用した各推定スペクトル $Y^*$  に対して音声区間検出関数を作用させて復元した目的音声信号を示す。図3と図10を比較すると、図10には雑音信号が残留していないことが判る。

【0078】

【数27】

$$Z(t) = \frac{1}{2\pi} \frac{1}{W(t)} \sum_k \sum_{\omega} e^{\sqrt{-1}\omega(t-k\tau)} Z(\omega, k)$$

$$W(t) = \sum_k \omega(t-k\tau) \dots\dots (27)$$

【0079】

本発明の第2の実施の形態に係る定常雑音下における音声区間検出に基づく目的音声の復元方法は、図1に示すように、異なる2つの音源11、12からそれぞれ発信される信号 $s_1(t)$ 及び信号 $s_2(t)$ （一方が目的音声信号であり、他方が雑音信号である）を、異なる位置に設けた第1及び第2のマイク13、14でそれぞれ受信して混合信号 $x_1(t)$ 、 $x_2(t)$ を形成し、各混合信号 $x_1(t)$ 、 $x_2(t)$ を時間領域から周波数領域にフーリエ変換して、独立成分分析法の一例であるFastICA法により目的音声及び雑音にそれぞれ対応する複数の推定スペクトル $Y^*$ 、 $Y$ を抽出する第1工程を有している。

また、目的音声の復元方法は、各推定スペクトル $Y^*$ を、各推定スペクトル $Y^*$ の振幅分布の尖度の値を分離判定基準にして雑音が除去された複数の推定スペクトル $y^*$ 及び雑音が残留する複数の推定スペクトル $y$ に分離する第2工程と、各推定スペクトル $y^*$ の総和 $F$ を求め、総和 $F$ の最大値に応じて決定される閾値 $\beta$ を検出判定基準にして総和 $F$ の時間領域から音声区間及び雑音区間をそれぞれ検出する第3工程を有している。

更に、目的音声の復元方法は、各推定スペクトル $Y^*$ を周波数領域から時間領域にフーリエ逆変換して得られた前記目的音声の信号から検出した音声区間に存在する成分を抽出して目的音声を復元する第4工程を有する。

【0080】

ここで、本発明の第2の実施の形態に係る定常雑音下における音声区間検出に基づく目的音声の復元方法は、第1の実施の形態の目的音声の復元方法と比較し、音声区間を時間領域で検出し目的音声の時間信号からこの音声区間に存在する成分のみを抽出して目的音声を復元することが大きな特徴となっている。

このため、第2の実施の形態に係る定常雑音下における音声区間検出に基づく目的音声の復元方法は、第1の実施の形態の目的音声の復元方法と比較して、第3及び第4工程が異なる。従って、第3及び第4工程についてのみ説明する。

【0081】

総和 $F$ におけるフレーム番号 $k$ とサンプリング時刻 $t$ との間には、フレーム周期を $\tau$ として、 $\tau(k-1) < t \leq \tau k$ の関係が成立している。従って、 $k = [t/\tau]$ の関係が成立していることを考慮すると、時間領域での音声区間検出関数 $F^*(t)$ を設定することができる。

すなわち、 $F^*(t)$ は、 $F^*([t/\tau]) = 1$ の領域で $F^*(t) = 1$ 、 $F^*([t/\tau]) = 0$ の領域で $F^*(t) = 0$ となる2値化関数として求まる。従って、 $F^*([t/\tau]) = 1$ となる $t$ の領域から音声区間を検出し、 $F^*([t/\tau]) = 0$ となる $t$ の領域から雑音区間をそれぞれ検出する（第3工程）。なお、 $[t/\tau]$ は、 $t/\tau$ 以上の最小の整数を指すceiling記号である。

また、時間領域での音声区間検出関数 $F^*(t)$ を、各推定スペクトル $Y^*$ を周波数領域から時間領域にフーリエ逆変換して得られた目的音声の時間信号に掛けることにより、雑音区間がゼロになって音声区間のみの音声が強調された信号を得ることができ（第4工程）、これを復元信号増幅器18を介してスピーカ19に入力することにより目的音声を確

認することができる。

【実施例】

【0082】

(実施例1)

図11に示すように、縦10m、横10m、及び高さ10mのバーチャルルーム内に音源1、2とマイク1、2を配置し、マイク1、2で観測される混合信号に対してFast ICA法を適用して雑音除去を行って目的音声を復元し、音声区間の検出精度を求めた。ここで、各マイク1、2間の距離は0.5m、各音源1、2間の距離は0.5m、床から各マイク1、2までの高さは1m、床から各音源1、2までの高さは1m、マイク1と音源1までの距離は0.5m、マイク2と音源2までの距離は0.5mである。またFast ICA法には、ゴタンダ、ノブ、コヤ、カネダ、イシバシ (H. Gotannda, K. Nobu, T. Koya, K. Kaneda, T. Ishibashi)、“パーミュテーション コレクション アンド スピーチ イクストラクション ベイスド オン スプリット スペクトラム スルー ファスト アイ シー エイ (Permutation Correction and Speech Extraction Based on Split Spectrum Through Fast ICA)”、プロシーディング オブ インターナショナル シンポジウム オン インディペンデント コンポーネント アナリシス アンド ブラインド シグナル セパレーション (Proc. International Symposium on Independent aration)、2003年4月1日、p379-384に記載した方法を採用した。

【0083】

音源1から目的音声として、「音声情報処理研究用 日本語音声データベース (平成3年度文部科学研究費補助金試験研究、研究代表者 板橋 秀一)」から5単語 (材料、いよいよ、羨ましい、面白い、具合) の男女各1名による発話パターン10種類を発信し、音源2から定常雑音としてノイズエクスター92 データベース (Noise x-92 database) [online] (<http://spib.rice.edu/spib/>) 中の5種類 (f16ノイズ (f16 noise)、ボルボノイズ (volvo noise)、ホワイトノイズ (white noise)、ピンクノイズ (pink noise)、タンクノイズ (tank noise)) を発信し、マットラブ コード (Matlab code) [online] (<http://sound.media.mit.edu/ica-bench/>) により50組の混合信号を生成した。

【0084】

音声区間検出関数 $F^*(k)$ は、総和 $F$ を閾値 $\beta$ で2値化することで生成し、総和 $F$ は推定スペクトル $Y^*$ を閾値 $\alpha$ で分離することにより生成している。音声区間の抽出精度は各閾値 $\alpha$ 、 $\beta$ に依存する。そのため、事前に最適な閾値 $\alpha$ 、 $\beta$ の値を調べた。その結果、閾値 $\alpha$ の値は1.8~2.3、閾値 $\beta$ の値は0.05~0.15とそれぞれ決定された。そこで、閾値 $\alpha=2.0$ 、閾値 $\beta=0.08$ とした。

混合信号から目的音声に対応する推定スペクトル $Y^*$ を抽出し、本発明の方法で音声区間を検出し始端と終端を求めた。また、推定スペクトル $Y^*$ から復元した目的音声の波形から目視により音声区間を検出し始端と終端を求めた。

その結果、本発明の方法により検出した音声区間の始端時刻は目視により検出した音声区間の始端時刻に対して-2.71msec (標準偏差13.49ms)、本発明の方法により検出した音声区間の終端時刻は目視により検出した音声区間の終端時刻に対して-4.96msec (標準偏差26.07msec) となって、音声区間を速めに検出する傾向を示したがその差は小さくほぼ正確に検出していることが判る。

【0085】

(実施例2)

音源2から非定常雑音としてNTT騒音データベース (NTTアドバンステクノロジー株式会社、アンビエント ノイズ データベース フォア テレホノメトリ (Ambient Noise Database for Telephonometry) 1996、N

TTアドバンステクノロジー株式会社、1996年)中の5種類の非定常雑音(オフィス(office)、レストラン(restaurant)、クラシカル(classical)、ステーション(station)、ストリート(street))を発信し、実施例1と同様の実験を行った。

その結果、本発明の方法により検出した音声区間の始端時刻は目視により検出した音声区間の始端時刻に対して-2.36msec(標準偏差14.12msec)、本発明の方法により検出した音声区間の終端時刻は目視により検出した音声区間の終端時刻に対して-13.40msec(標準偏差44.12msec)となつて、雑音が非定常性を帯びている場合でも目視法と同程度の機能を有し、音声区間をほぼ正確に検出していることが判る。

#### 【0086】

以上、本発明の実施の形態を説明したが、本発明は、この実施の形態に限定されるものではなく、発明の要旨を変更しない範囲での変更は可能であり、前記したそれぞれの実施の形態や変形例の一部又は全部を組み合わせる場合も本発明の権利範囲に含まれる。

例えば、目的音声及び雑音にそれぞれ対応する複数の推定スペクトル $Y^*$ 、 $Y$ を抽出するのにFastICA法を適用したが、ICA法で分離した各スペクトルに対して各マイクと各音源との距離に基づく音の伝達特性を考慮してスケージングの不定性を解決すると共に、各周波数で各スペクトルの包絡を求めて、その類似度に基づく成分置換の解消を行つて推定スペクトル $Y^*$ 、 $Y$ を抽出することもでき、推定スペクトル $Y^*$ 、 $Y$ の抽出方法に指定はない。

#### 【図面の簡単な説明】

##### 【0087】

【図1】本発明の第1及び第2の実施の形態に係る定常雑音下における音声区間検出に基づく目的音声の復元方法を適用した目的音声の復元装置の構成図である。

【図2】同復元方法における目的音声及び雑音から各推定スペクトル $Y^*$ が形成されるまでの信号の流れを示す説明図である。

【図3】各推定スペクトル $Y^*$ から構成された復元スペクトル群を周波数領域から時間領域にフーリエ逆変換して得られた目的音声信号の波形を示す説明図である。

【図4】雑音が除去された推定スペクトル $Y^*$ の波形を示す説明図である。

【図5】雑音が残留している推定スペクトル $Y^*$ の波形を示す説明図である。

【図6】雑音が除去された推定スペクトル $Y^*$ の振幅分布を示す説明図である。

【図7】雑音が残留している推定スペクトル $Y^*$ の振幅分布を示す説明図である。

【図8】雑音が除去された各推定スペクトル $y^*$ の総和を示す説明図である。

【図9】各推定スペクトル $Y^*$ に対して決定された音声区間検出関数のグラフである。

。【図10】各推定スペクトル $Y^*$ から音声区間に存在する成分を抽出して生成した目的音声の復元スペクトル群を周波数領域から時間領域にフーリエ逆変換して得られた目的音声信号の説明図である。

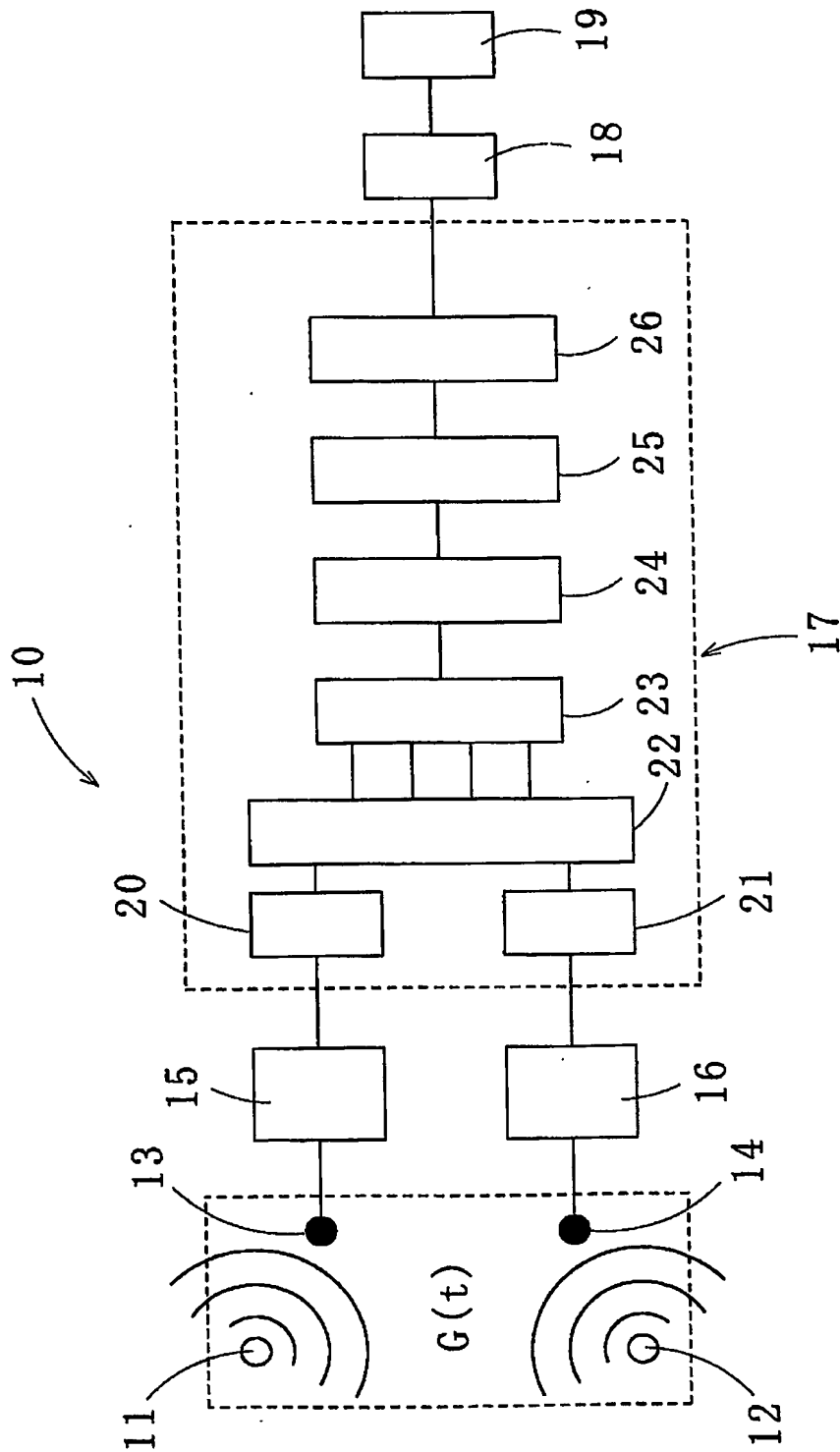
【図11】本発明の実施例1、2におけるバーチャルルームでの音源とマイクの配置を示す説明図である。

#### 【符号の説明】

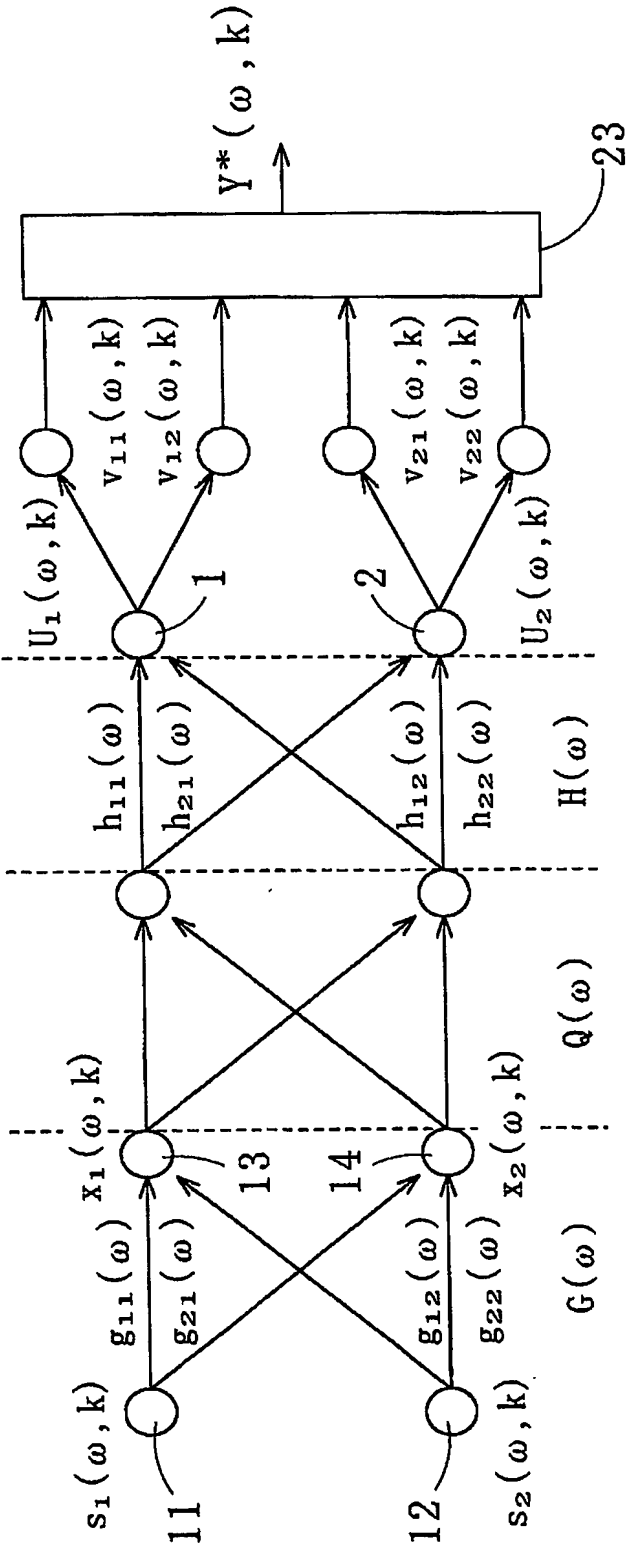
##### 【0088】

10:目的音声の復元装置、11、12:音源、13:第1のマイク、14:第2のマイク、15:第1の増幅器、16:第2の増幅器、17:復元装置本体、18:復元信号増幅器、19:スピーカ、20、21:A/D変換器、22:分割スペクトル生成器、23:推定スペクトル抽出回路、24:音声区間検出回路、25:復元スペクトル抽出回路、26:復元信号生成回路

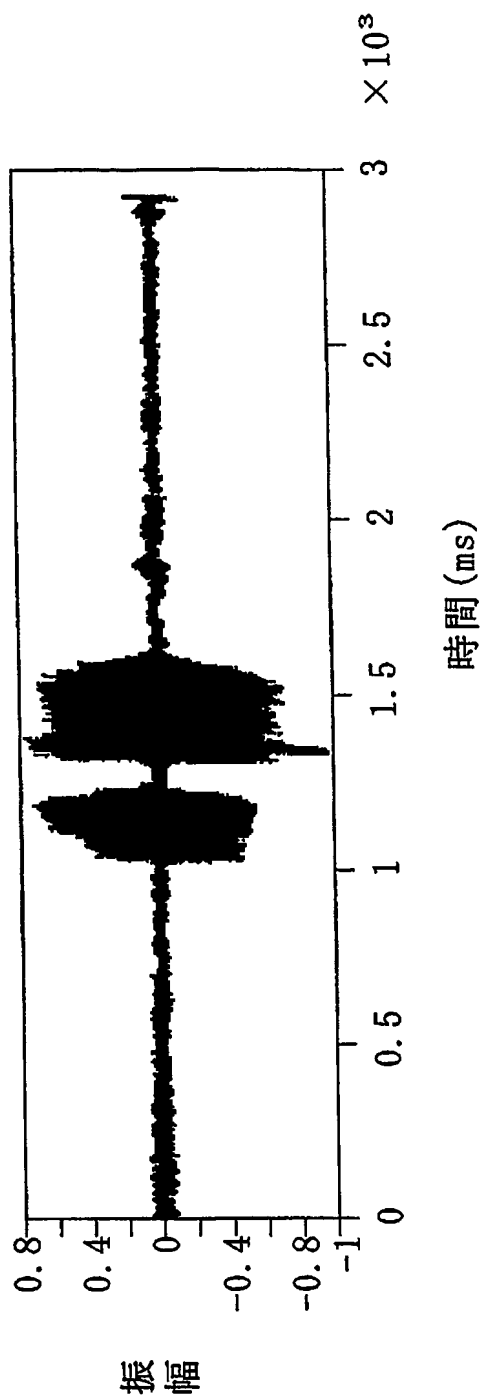
【書類名】 図面  
【図 1】



【図 2】

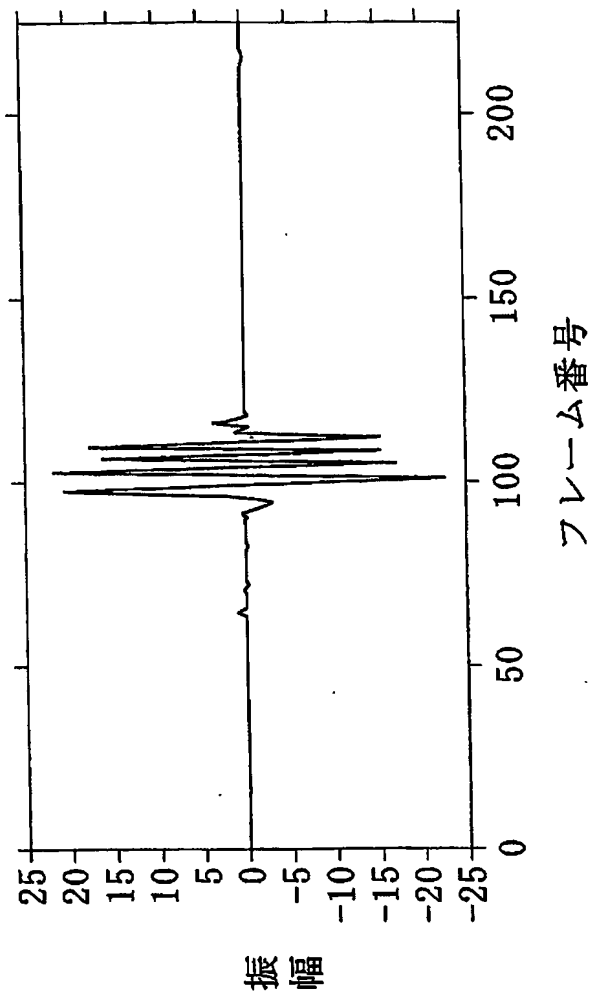


【図 3】

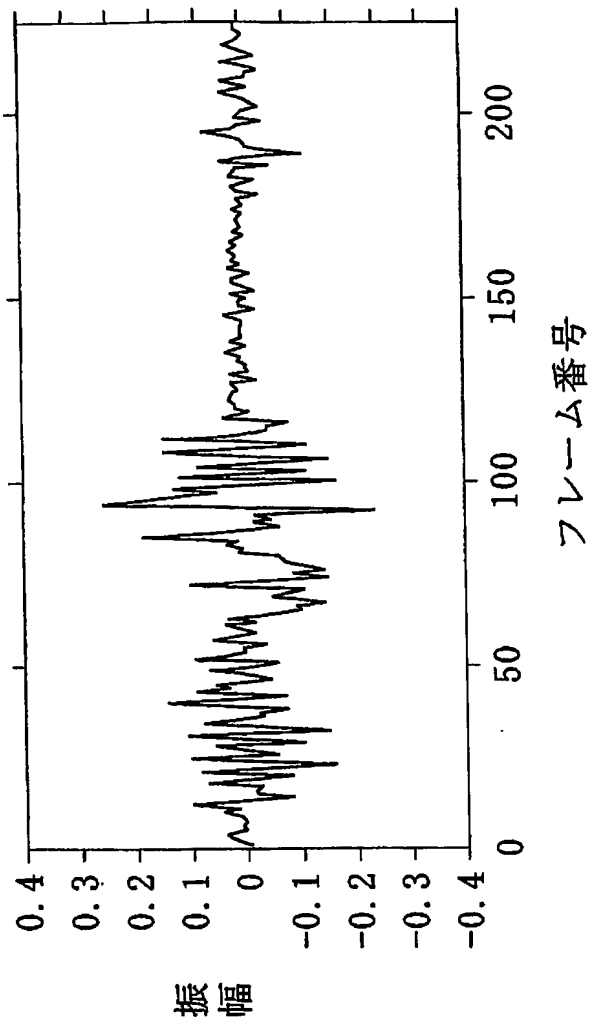




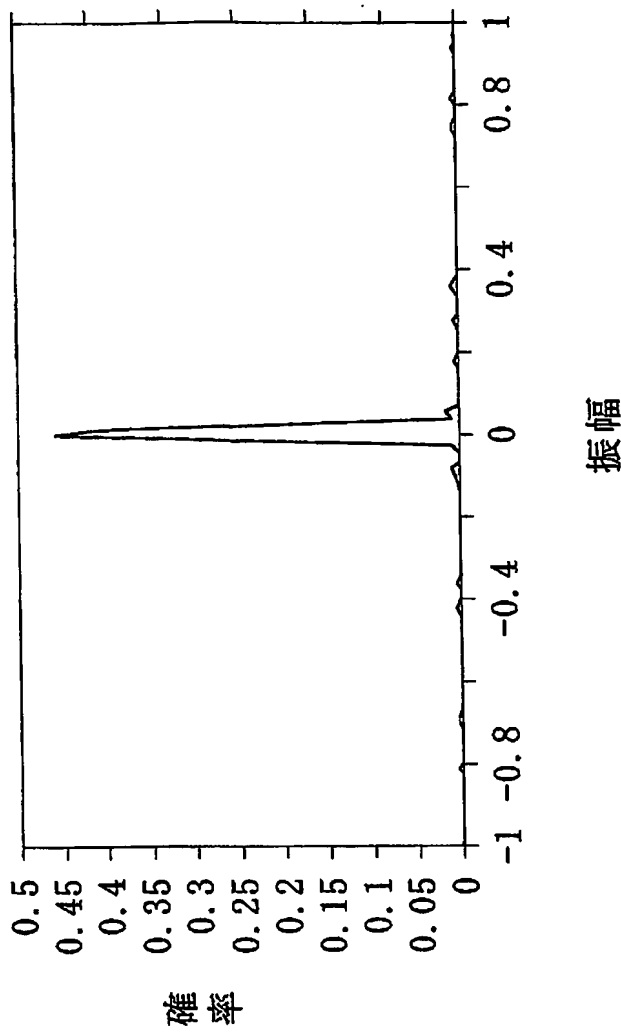
【図 4】



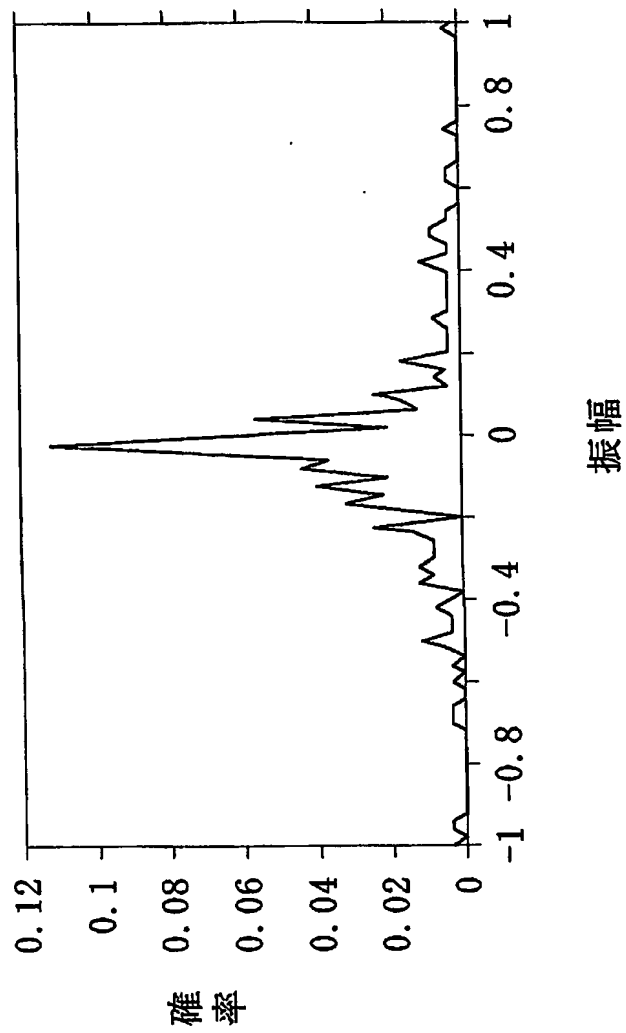
【図 5】



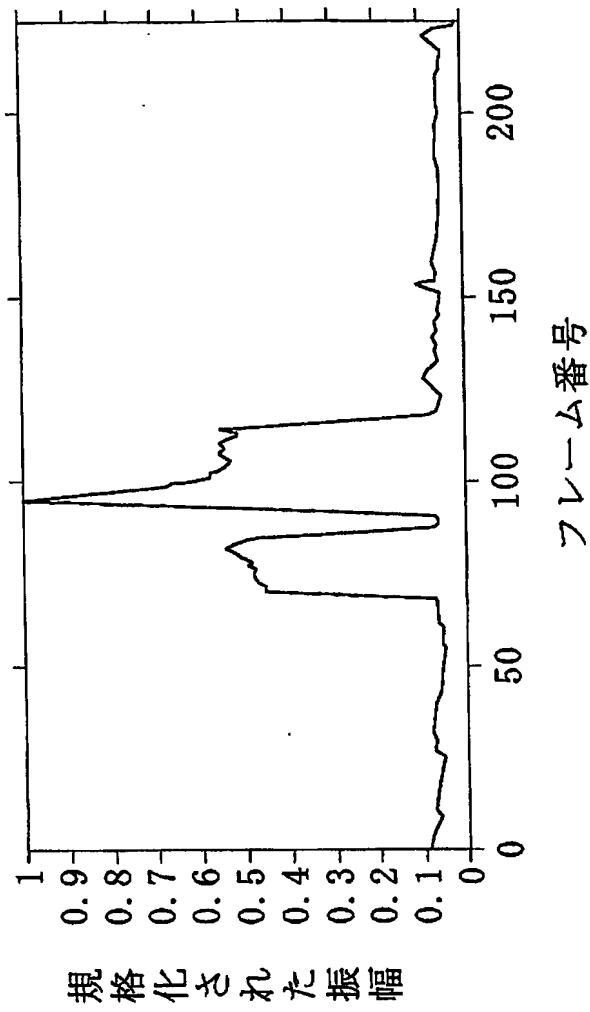
【図 6】



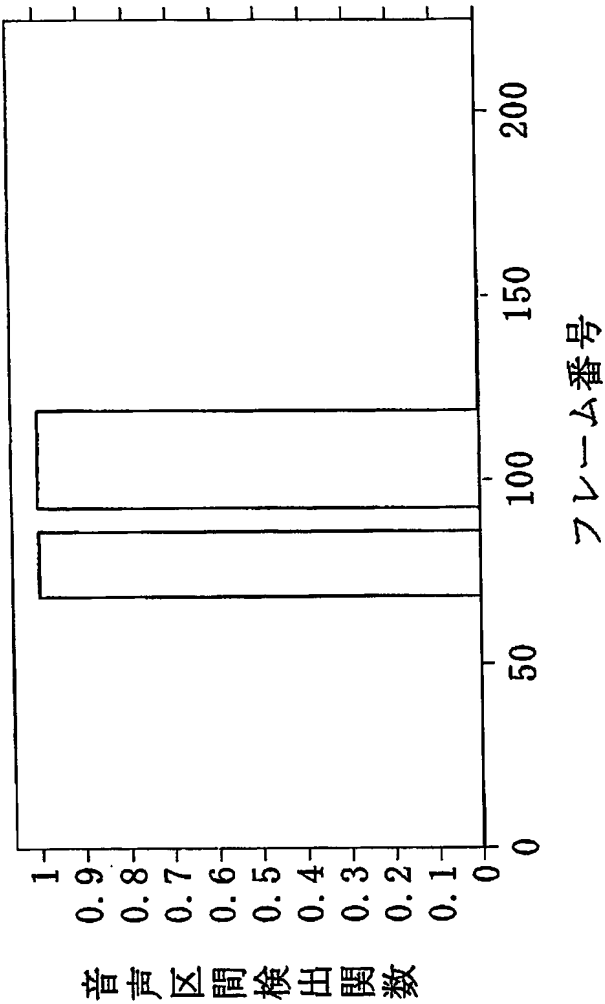
【図 7】



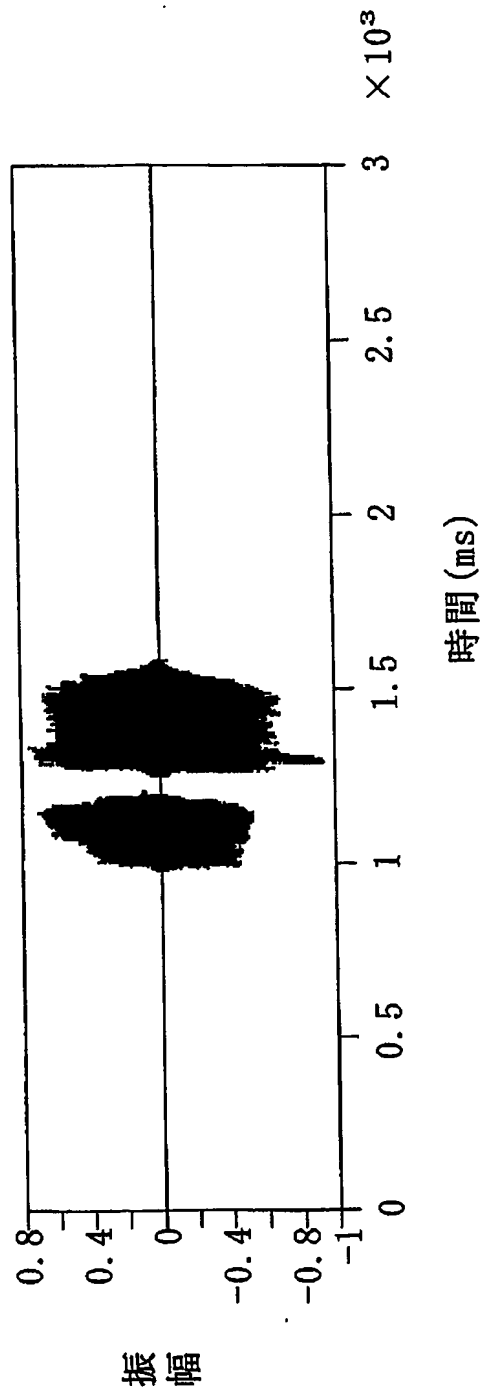
【図 8】



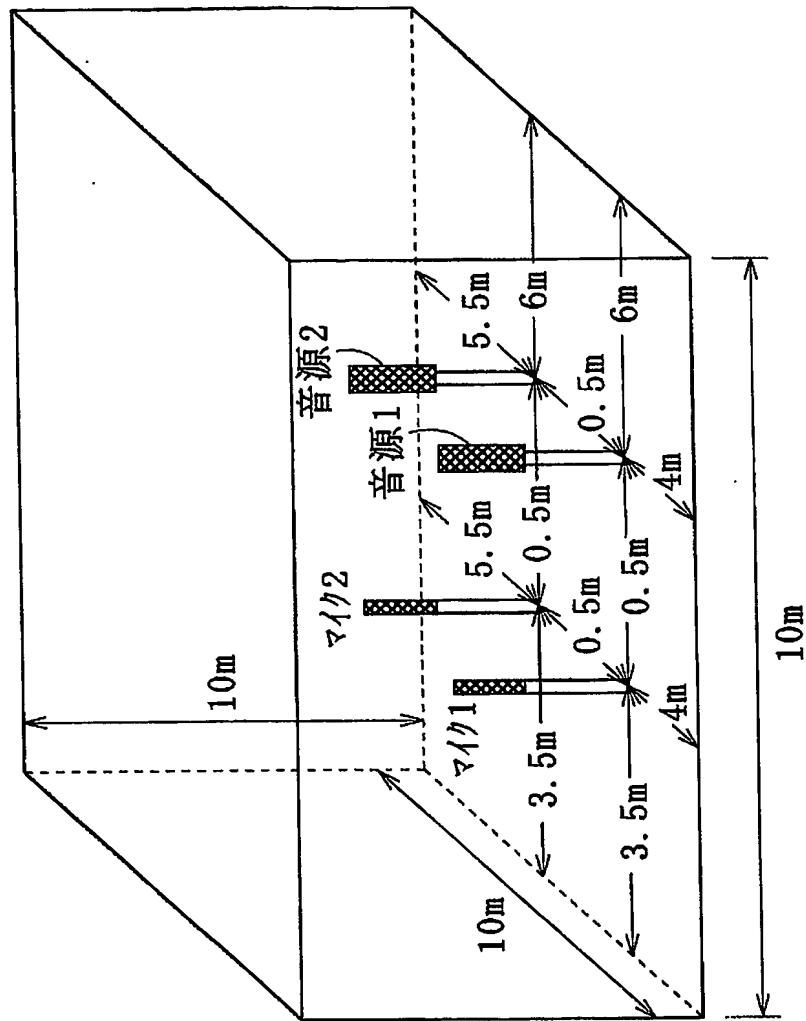
【図9】



【図 10】



【図 11】





## 【書類名】 要約書

## 【要約】

【課題】 ICA法により作成した分離信号から音声区間に含まれる信号のみを抽出して残留雑音を最小化することが可能な目的音声の復元方法を提供する。

【解決手段】 2つの音源11、12からそれぞれ発信される目的音声及び雑音を、第1及び第2のマイク13、14でそれぞれ受信し独立成分分析法により目的音声に対応する推定スペクトル $Y^*$ を抽出する第1工程と、推定スペクトル $Y^*$ からその振幅分布の尖度の値を分離判定基準にして雑音が除去された推定スペクトル $y^*$ を分離する第2工程と、推定スペクトル $y^*$ の総和 $F$ を求め、その最大値に応じて決定される閾値 $\beta$ を検出判定基準にして総和 $F$ の周波数領域から音声区間及び雑音区間をそれぞれ検出する第3工程と、推定スペクトル $Y^*$ から音声区間に存在する成分を抽出して目的音声の復元スペクトル群を生成し目的音声を復元する第4工程とを有する。

【選択図】 図1

認定・付加情報

特許出願の番号	特願 2003-314247
受付番号	50301477018
書類名	特許願
担当官	古田島 千恵子 7288
作成日	平成15年 9月 8日

<認定情報・付加情報>

【提出日】 平成15年 9月 5日

特願 2 0 0 3 - 3 1 4 2 4 7

出 願 人 履 歴 情 報

識別番号 [ 0 0 0 1 2 5 3 4 7 ]

1. 変更年月日	1 9 9 0 年 8 月 8 日
[変更理由]	新規登録
住 所	大阪府東大阪市小若江 3 丁目 4 番 1 号
氏 名	学校法人近畿大学

特願 2003-314247

出 願 人 履 歴 情 報

識別番号 [802000031]

1. 変更年月日 2002年 4月19日

[変更理由] 新規登録

住 所 福岡県北九州市若松区ひびきの2番1号

氏 名 財団法人北九州産業学術推進機構

特願 2 0 0 3 - 3 1 4 2 4 7

出 願 人 履 歴 情 報

識別番号 [ 5 0 0 3 3 1 2 1 9 ]

1. 変更年月日 2 0 0 2 年 7 月 2 日

[変更理由] 住所変更

住 所 福岡県福岡市博多区博多駅東一丁目 1 7 番 2 5 号

氏 名 株式会社ウェイブコム

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☒ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**